Received: 30 January 2024 | Revised: 5 July 2024 | Accepted: 8 July 2024

DOI: 10.1002/mp.17312

RESEARCH ARTICLE

MEDICAL PHYSICS

Deep learning-based dose prediction for magnetic resonance-guided prostate radiotherapy

Samuel Fransson[1,2] | **Robin Strand**[2,3] | **David Tilly**[1,4]

[1]Department of Medical Physics, Uppsala University Hospital, Uppsala, Sweden

[2]Department of Surgical Sciences, Uppsala University, Uppsala, Sweden

[3]Department of Information Technology, Uppsala University, Uppsala, Sweden

[4]Department of Immunology, Genetics and Pathology, Uppsala University, Uppsala, Sweden

Correspondence
Samuel Fransson, Department of Medical Physics, Uppsala University Hospital, 75185 Uppsala, Sweden.
Email: samuel.fransson@uu.se

Funding information
ASSIST, Grant/Award Number: 2019-04735

Abstract

Background: Daily adaptive radiotherapy, as performed with the Elekta Unity MR-Linac, requires choosing between different adaptation methods, namely ATP (Adapt to Position) and ATS (Adapt to Shape), where the latter requires daily re-contouring to obtain a dose plan tailored to the daily anatomy. These steps are inherently resource-intensive, and quickly predicting the dose distribution and the dosimetric evaluation criteria while the patient is on the table could facilitate a fast selection of adaptation method and decrease the treatment times.

Purpose: In this work, we aimed to develop a deep-learning-based dose-prediction pipeline for prostate MR-Linac treatments.

Methods: Two hundred twelve MR-images, structure sets, and dose distributions from 35 prostate patients treated with 6.1 Gy for 7 or 6 fractions at our MR-Linac were included, split into train/test partitions of 152/60 images, respectively. A deep-learning segmentation network was trained to segment the CTV (prostate), bladder, and rectum. A second network was trained to predict the dose distribution based on manually delineated structures. At inference, the predicted segmentations acted as input to the dose prediction network, and the predicted dose was compared to the true (optimized in the treatment planning system) dose distribution.

Results: Median DSC values from the segmentation network were 0.90/0.94/0.87 for CTV/bladder/rectum. Predicted segmentations as input to the dose prediction resulted in mean differences between predicted and true doses of 0.7%/0.7%/1.7% (relative to the prescription dose) for $D_{98\%}/D_{95\%}/D_{2\%}$ for the CTV. For the bladder, the difference was 0.7%/0.3% for $D_{mean}/D_{2\%}$ and for the rectum 0.1/0.2/0.2 pp (percentage points) for $V_{33Gy}/V_{38Gy}/V_{41Gy}$. In comparison, true segmentations as input resulted in differences of 1.1%/0.9%/1.6% for CTV, 0.5%/0.4% for bladder, and 0.7/0.4/0.3 pp for the rectum. Only $D_{2\%}$ for CTV and $D_{mean}/D_{2\%}$ for bladder were found to be statistically significantly better when using true structures instead of predicted structures as input to the dose prediction.

Conclusions: Small differences in the fulfillment of clinical dose-volume constraints are seen between utilizing deep-learning predicted structures as input to a dose prediction network and manual structures. Overall mean differences <2% indicate that the dose-prediction pipeline is useful as a decision support tool where differences are >2%.

KEYWORDS
deep learning, dose prediction, radiotherapy

1 | INTRODUCTION

Daily adaptive radiotherapy of prostate patients, as performed with an MR-Linac,[1,2] offers opportunities for improving treatments with accurate dose delivery and the possibility for margin reduction. However, the prolonged treatment times as compared to conventional external radiotherapy treatments induce obstacles for full adoption, where treatment times can stretch for up to or above 1 h per fraction (as observed in our clinic and with others[3]). On the Elekta Unity MR-Linac, the choice between the Adapt to Position (ATP) and the Adapt to Shape (ATS) workflow[1] has to be made. The latter method is more precise but much slower due to the requirement of re-contouring the structures, a task requiring 10–15 min for a prostate case at our institution. Additionally, the plan optimization may add 1–2 min more in comparison to ATP.[1] Although the ATS workflow is predominantly used at our institution, it is not always clear if this is the best choice. A retrospective analysis showed that up to 84% of prostate fractions could have been treated using the ATP workflow while still meeting the dose-volume constraints (DVCs) (see Supplementary Material). Additionally, the longer treatment times for ATS may lead to reduced patient throughput and greater intrafraction movement.[4,5] Therefore, efforts to support the ATP/ATS decision have been attempted, including, for example, deformable image registration (DIR) based methods,[6,7] decision tree analysis,[8] and machine learning classifier based on wavelets[9] to find factors correlating with plan quality.

Deep learning-based approaches are becoming common in radiotherapy, with clinical adoption of, for example, organ at risk (OAR) segmentation and sCT (synthetic Computed Tomography) generation. Segmentation systems may reduce the inter- and intra-observer variations of the treatment contours and offer substantial time savings,[10–12] while sCT generation can reduce the number of scans required.[13] Deep learning-based approaches could also be applied in the prediction of radiation dose. In the MR-Linac scenario, obtaining a fast and accurate prediction of the likely dose distribution could assist in the choice between ATS and ATP, affecting the overall treatment times. In addition, a dose prediction tool can be used to evaluate the dosimetric impact of, for example, choice of treatment margins and contouring strategies.[14,15] Several previous works focused on predicting the dose for radiotherapy prostate treatments. The seminal work of Nguyen et al.[16] utilized the ubiquitous U-net[17] architecture, and this has been extended to both 2.5D[18] and 3D[19–21] and can be modified to include attention gates.[22] Although most of the approaches are not aimed toward dose prediction for MR-based treatments, they rely heavily, or even exclusively, on the input of contours to be accurate, which also makes the method independent of imaging modality. Dose prediction has been performed with solely an image as input,[23] but with inferior accuracy compared to when the input included the contours. Simultaneous segmentation and dose prediction have been performed in a multitask approach,[24] concluding improved performance over only predicting the dose. In the context of MR-linac, dose prediction on contours of abdominal treatments was found useful for finding plans with inferior quality.[25]

For a decision support tool to be useful within the context of an MR-Linac treatment the prediction must precede any form of manual re-contouring in order to decrease the treatment time. However, any evaluation based on DVCs relies on the presence of contours. Therefore, the aim of this work was to combine a standalone segmentation network to produce contours, followed by a dose prediction network of the segmentation output to a potential decision support tool for prostate MR-Linac treatments. A pipeline that separates the prediction of contours and dose distribution could help by supporting the choice between ATS/ATP workflows and additionally offering a good starting point for the contours if choosing ATS, which can then be manually refined.

TABLE 1 Dose-volume constraints for the target and organs at risk.

PTV	CTV	Bladder	Rectum
$D_{98\%} > 95\%$	$D_{98\%} > 98\%$	$D_{mean} < 34$ Gy	$V_{33Gy} < 30\%$ $V_{38Gy} < 15\%$ $V_{41Gy} < 10\%$

2 | MATERIALS AND METHODS

We subdivided the methods for the dose prediction tool into (1) a segmentation network producing contours of Clinical Target Volume (CTV) (prostate), bladder, and rectum and (2) a dose prediction network taking the MR-image and contours as input to predict a dose distribution.

2.1 | Dataset

The dataset consisted of 212 MR-images, contours, and dose distributions from 35 prostate patients treated with the ATS workflow with 6.1 Gy × 7 or 6 fractions at our MR-Linac between 2019 and 2023. The use of this data was approved by the Swedish Ethics Review Authority (Approval ID: 2019–03050). The treatment plan was a 7-field step-and-shoot IMRT plan optimized to adhere to the DVCs (see Table 1).

In the experiments, we additionally added $D_{2\%}$ for the target volumes and bladder to increase the sensitivity in evaluating the dose distributions.

The images are T2-weighted 3D images (turbo spin echo acquisition with 90° flip angle and average repetition/echo times of 1513/263 ms) with a reconstructed voxel size of $0.86 \times 0.86 \times 1$ mm^3 (left-right, anterior-posterior, superior-inferior). The resolution of the dose matrices was $3 \times 3 \times 3$ mm^3. Due to rapid contouring online, only the CTV and the part of the OARs in its vicinity were accurately contoured, and performed by different physicians. Accurate contouring of OARs (bladder and rectum) was performed retrospectively, by a single observer. Other available contours were the PTV, body contour, and bone structure (all bones within the body contour as a single structure). Our institution employs an MR-only workflow where electron densities are assigned to the body contours and the bones for the sake of dose calculation. In general, the CTV to PTV margins were 7 mm, except for 5 mm posteriorly. There were exceptions for one patient with a 5 mm isotropic margin, one patient with 6 mm except for 5 mm posteriorly, and two patients with 5 mm except for 3 mm posteriorly.

Due to the optimization and thereby the obtained dose distributions depending on the outline of the contours at the online planning occasion, measures were taken to evaluate and minimize the possible dose differences between optimizing with online OARs as opposed to with retrospectively corrected OARs (see Figure S1). In short, the fractions exhibiting the largest difference between corrected and online OARs in terms of Dice Similarity Coefficient (DSC)[26] were selected. A dose optimization was then performed with the corrected contours, and the obtained dose distribution was compared to the online dose distribution.

Since all initial data was in DICOM image format, it was converted to Python numpy arrays utilizing dcmrtstruct2nii[27] and SimpleITK.[28–30] Before training, all datasets were cropped around the body contour to alleviate the computational burden and to decrease the large amount of background voxels. Additionally, the images were intensity z-normalized by subtracting the intensity mean value and dividing by the standard deviation, and the dose distributions were normalized with the prescribed dose. The images (and corresponding contours and dose distributions) were subdivided into a training and a testing cohort, consisting of 152 images (25 patients) and 60 images (10 patients), respectively.

2.2 | Deep learning image segmentation

We utilized a 3D implementation of the U-net[17] with added attention gates.[31] Other hyperparameters and training settings were inspired by the nnU-net[32] due to its high performance across a large variety of anatomies for segmentation. We used 30 initial layers, three max-pooling layers, instance normalization,[33] Leaky ReLU (slope 1e-2) in the hidden layers and a softmax activation at the last layer. We utilized the same loss function as in the nnU-net, combining cross-entropy and soft dice loss with equal contributions. All training was performed with Keras and Tensorflow 2.4 as backend on an Nvidia RTX 3090 GPU. Due to memory constraints on the GPU, training was performed with randomly extracted patches of $128 \times 128 \times 128$ and a batch size of 2. A 5-fold cross-validation training scheme was performed with the training data where each model was trained for 500 epochs to ensure proper convergence. For each fold, all images from 20 patients were used as training data, and images from the remaining five as validation. Here, an epoch is defined as when all images have supplied one patch to the update of the network weights. The Adam optimizer[34] with learning rate 1e-4 was applied. Image augmentation by randomly applying left-right flipping was employed. At inference, a patch-based approach was employed with a patch size of $192 \times 192 \times 192$ with a stride over the images, overlapping with 96 voxels and weighted by a Hann window.[35] When models were applied to the test images, an ensemble approach took the average of each model's softmax output to achieve the final prediction. Since all structures should be volumetrically coherent, a postprocessing step of keeping only the largest coherent volume for each contour was applied.

2.3 | Deep learning dose prediction

Similar to the segmentation network, a 3D U-net with attention gates was employed for the dose prediction,[22] but with a linear activation in the final layer. Also, for the upsampling step in the decoder, a nearest neighbor-based approach was utilized since transposed convolution yielded significant checkerboard artifacts in the final prediction. The input to the network was the image, PTV, bladder, rectum, bone structure, and body contour, concatenated along the channel (last) dimension. The training data for this network was the corrected manually segmented contours, that is, not those obtained from the segmentation network. The loss function utilized was a combination of mean squared error (MSE) and a moment-based loss function,[36] chosen for its simplicity and high performance as compared to other DVH-based loss functions. We utilized the same moments as in the original publication but increased its relative weight to 0.1 since it gave slightly better results on the validation data. Here, a patch size of $64 \times 64 \times 64$ was utilized during training and a 5-fold cross-validation was performed with the same data split as from the training of the segmentation network. During inference, the same patch size was employed with an overlap of 32 voxels. When applied to the test images, the average of the output of all five models was taken to obtain the final predicted dose.

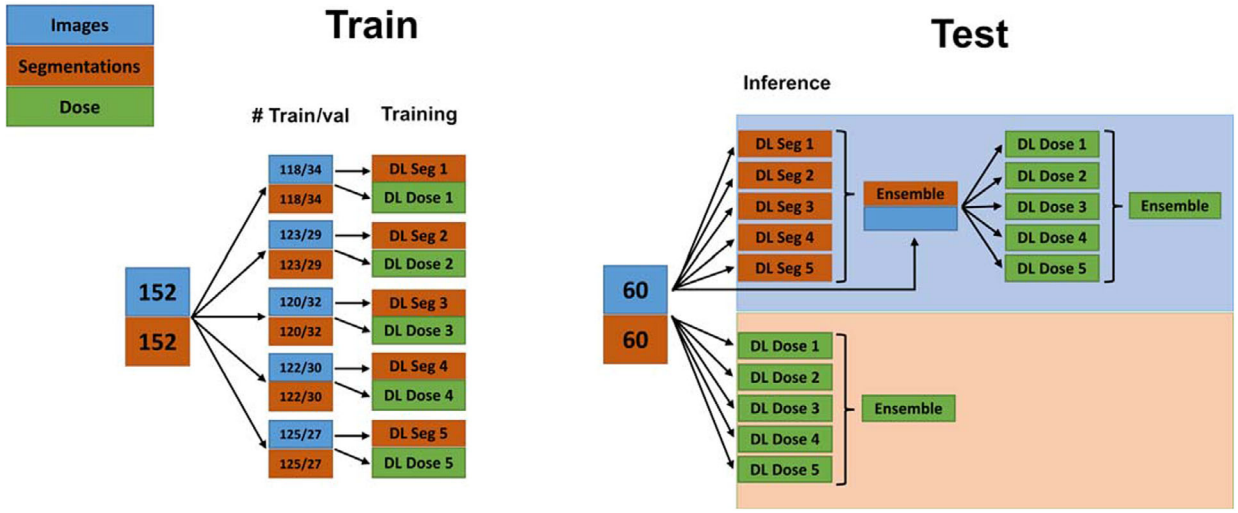

FIGURE 1 Outline of the training and inference procedure. The training is performed with a 5-fold cross-validation creating five segmentation and dose prediction models, respectively. In the testing phase, an ensemble method is applied by taking the average of the model outputs. When applied to the test data, both predicted structures (upper part) and true structures (lower part) are set as input to the dose prediction models to allow for comparison.

2.4 | Workflow inference

The inference of the full dose prediction workflow consisted of initially performing image segmentation to obtain the CTV, bladder, and rectum. A PTV was generated from the CTV according to the margins from the actual treatment. The body contour and bone structure were taken from the treatment data. The resulting array acted as input to the dose prediction network, and the following predicted dose distribution was rescaled with the prescription dose. The workflow for both training and inference is shown in Figure 1.

3 | RESULTS

3.1 | Deep learning image segmentation

The segmentation results are shown in Figure 2 for the test data set, evaluated with DSC, Hausdorff 95%, and the volume difference between the true and predicted structures. The volume differences were added since the structure volumes are important for the dose prediction evaluation. The PTV is not predicted by the segmentation network but was generated based on margins around the CTV. Similar results for the validation data are available in Figure S2.

3.2 | Deep learning dose prediction

The dose prediction results in terms of DVC are shown in Figure 3 for the test data set. Similar results for the validation data are available in Figure S3. The input to the dose prediction was both the true and the predicted contours, to determine the influence of the segmentation network. However, all doses were evaluated on the true contours for a relevant comparison. For CTV/PTV, $D_{2\%}$ was consistently underestimated with a mean value of 1.6%/1.5% and 1.7%/1.5% (relative to prescription dose) for true and predicted input contours, respectively. With predicted contours as input, a clear trend for the PTV was an underestimation with a mean deviation of 3.2% and 1.6% for $D_{98\%}$ and $D_{95\%}$, respectively. For OARs, the difference between the true and predicted contour input was small, and the true contour input was found significantly better from a one-sided Wilcoxon signed rank test only for the bladder. The mean deviation from the true values for bladder were 0.7%/0.3% for $D_{mean}/D_{2\%}$ and for rectum 0.1/0.2/0.2 pp (percentage points) for $V_{33Gy}/V_{38Gy}/V_{41Gy}$. To further evaluate the effect, we calculated for how many instances the true and predicted dose values lie on different sides of the threshold, for example, constraint fulfilled on true dose distribution but not on predicted and vice versa. This was the case for 5/60 of the test data instances, all due to at least one of the rectum constraints. Of these, four were violations of the clinical dose distribution and one of the predicted dose distribution.

In Figure 4, a dose-volume histogram for the test data is shown. The histogram is averaged over all instances in the test dataset, hence depicting the average values for all patients, and the evaluation performed on the true contours.

In Figure 5 example results are shown from two representative cases with respect to the PTV coverage (upper row) and adherence to rectum constraints (bottom row).

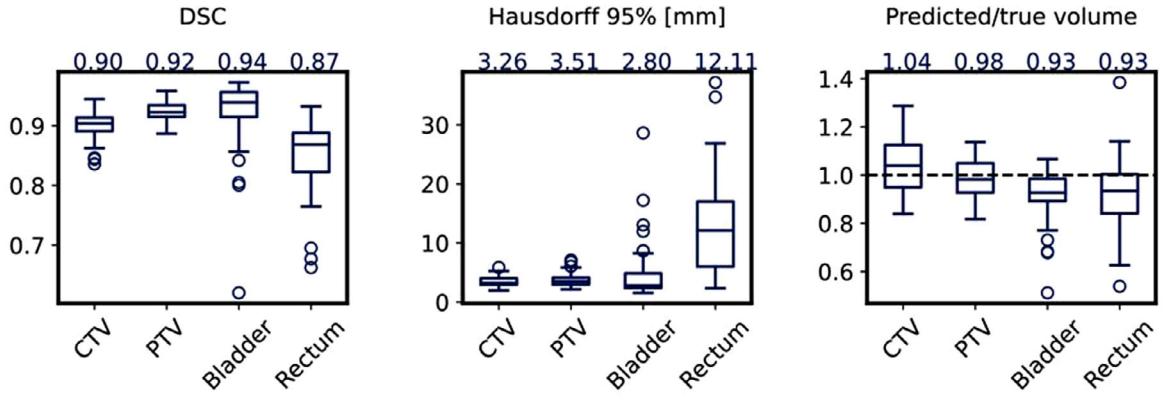

FIGURE 2 Results from the segmentation network. Plots from left to right correspond to the evaluation metrics DSC, Hausdorff 95%, and volume differences, respectively. For each plot, the results for CTV, PTV, bladder, and rectum are shown for the test data consisting of 60 data points from 10 patients. Numbers above each plot indicate the median value. The upper whisker is set to extend to the largest datapoint smaller than Q3 + 1.5*IQR and the lower whisker to the smallest datapoint larger than Q1 - 1.5*IQR, where Q3 and Q1 are the upper and lower quartiles and IQR the interquartile range (Q3-Q1). Points outside are considered outliers and denoted with circles.

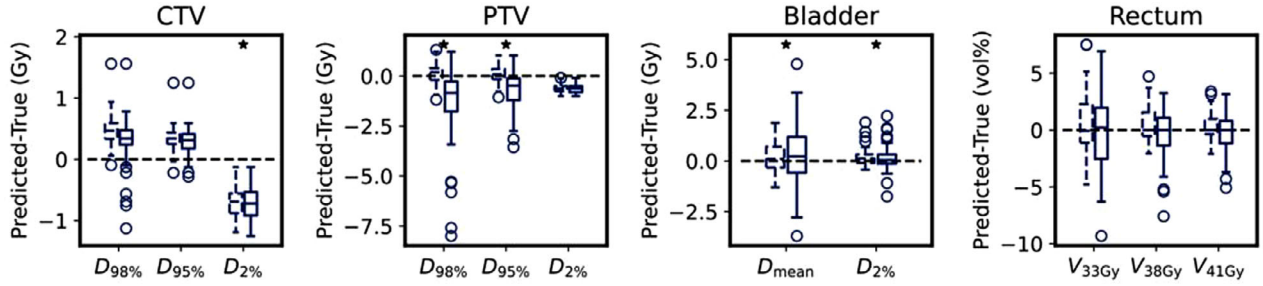

FIGURE 3 The difference in dose-volume constraint between the true and predicted dose for CTV, PTV, bladder, and rectum, respectively, when the input to the dose prediction network was the true structures (dashed lines) and predicted structures (solid lines), evaluated on the test data set consisting of 60 data points from 10 patients. The difference was taken with respect to the full prescribed dose for the dose difference (CTV, PTV, bladder) and volume percentage (rectum). The rectum constraints, that is, V_{xx}, are given with xx in Gy. Stars above each pair of boxes indicate that dose prediction based on true structures is significantly better than on predicted structures, depicted with a one-sided Wilcoxon signed rank test with $p < 0.05$. The upper whisker is set to extend to the largest datapoint smaller than Q3+1.5*IQR and the lower whisker to the smallest datapoint larger than Q1-1.5*IQR, where Q3 and Q1 are the upper and lower quartiles and IQR the interquartile range (Q3-Q1). Points outside are considered outliers and denoted with circles.

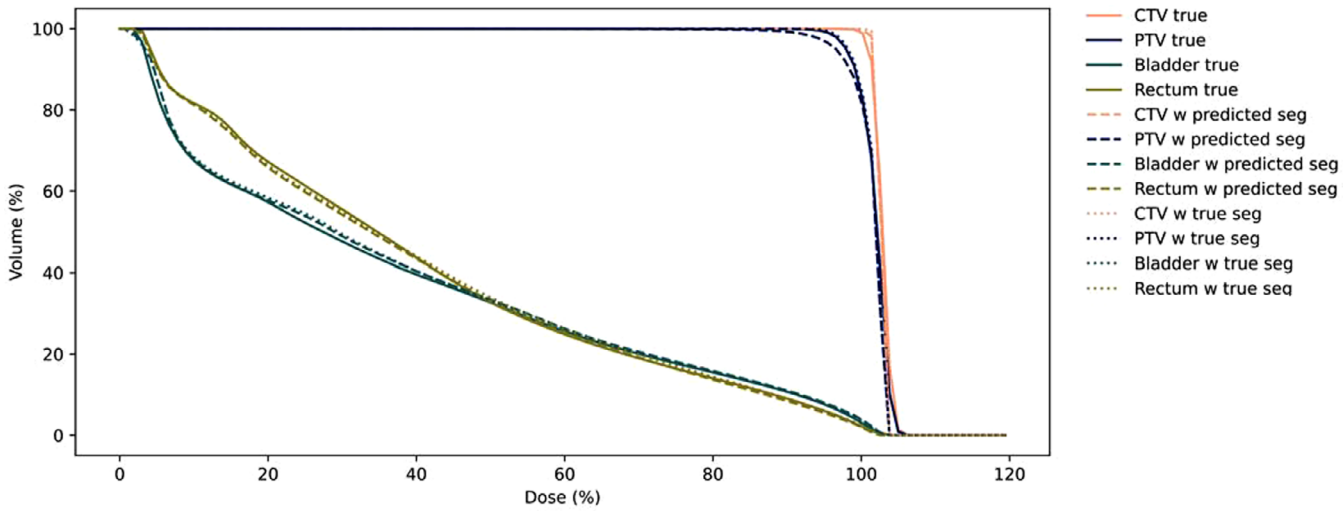

FIGURE 4 Average dose-volume histograms for the test dataset. Solid lines show the true distribution, dashed the predicted distribution with predicted structures as input, and dotted the predicted distribution with true structures as input. All distributions are evaluated on the true structures.

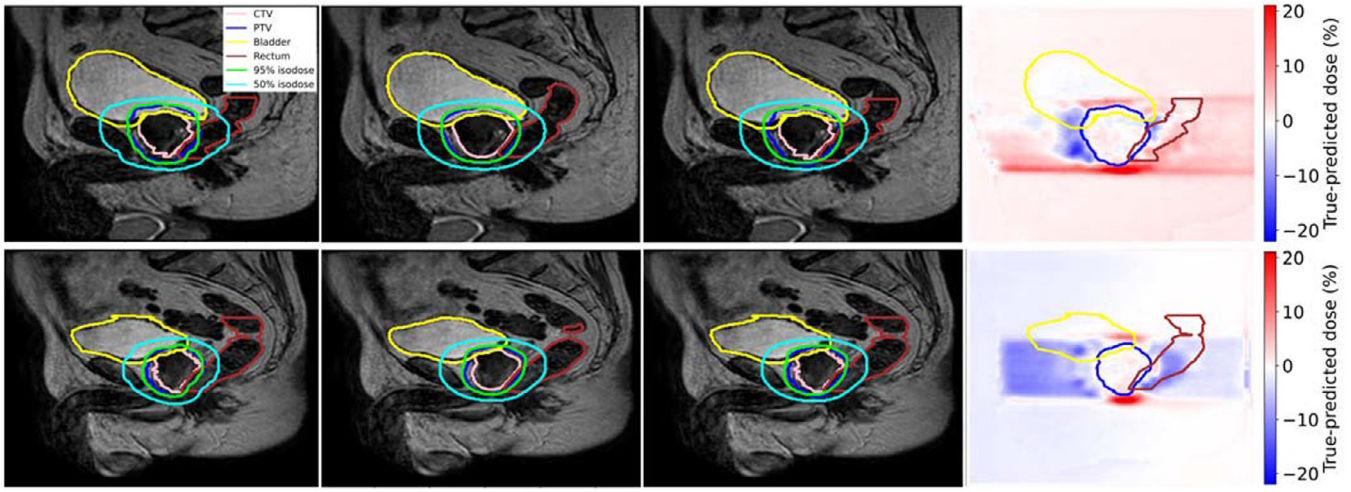

FIGURE 5 Example plots of two representative cases with respect to the PTV coverage (upper row) and adherence to rectum constraints (bottom row). CTV, PTV, bladder, and rectum are shown in pink, blue, yellow, and red, respectively. Isodose lines of 95% and 50% are shown with green and light blue, respectively. From left to right are (1) the true dose superimposed on the true structures, (2) the predicted dose superimposed on the predicted structures, (3) the predicted dose superimposed on the true structures, and (4) dose differences (true-predicted based on predicted structures) superimposed on the true structures.

3.3 | Inference times

The mean inference times for the full pipeline from image to predicted dose was 110 [range 70–170] seconds for the images in the test dataset. This includes the segmentation, dose prediction, and all pre- and post-processing steps required such as resampling and normalization. The mean segmentation time was 101 [range 60–162] seconds and the mean dose prediction time 4 [range 3–7] seconds.

4 | DISCUSSION

In this work, we have developed a complete pipeline for deep learning-based dose prediction for MRL prostate treatments from images to segmentation and dose distribution. It was divided into two separate steps, segmentation of relevant contours followed by a dose prediction based on the predicted contours. Mean differences <1%/1pp for all DVCs for bladder/rectum and <2% (with respect to prescription dose) for CTV indicate that the dose-prediction pipeline is useful as a decision support tool where differences are >2%. In addition, high overall median DSC values ≥0.87 indicate the usability of the predicted contours in an ATS workflow.

As for the segmentation results, the rectum exhibits the lowest value of 0.87, which is mainly due to the difficulty of determining the cranial-caudal extent of the contour. Especially in the cranial direction, the predicted segmentation did not have as clear a cut of the contour as the manual, as examplified in Figure 5. However, the accuracy of the segmentation should preferably be highest in the vicinity of the PTV since distal parts will receive only a small dose. Evaluating the rectum results only on axial slices containing PTV ± 1 cm resulted in an increased DSC to 0.89.

As for the dose prediction, $D_{2\%}$ for CTV/PTV is consistently underestimated, which may be due to the predicted dose being smoother than the true dose distribution calculated by the TPS. When comparing the dose prediction for true versus predicted contour input, the target coverage is also slightly underestimated for the latter. As seen in Figure 3, the PTV coverage based on the predicted CTV is on average ∼2% smaller than based on the true PTV. The PTV from the predicted CTV is based on an in-house Python implementation and resulted in volumes ∼1% smaller than those generated by the TPS, which is likely to only have a marginal influence on the predicted dose.

Although other works do not report all the same DVCs, some overlap can be found. Kandalan et al.[20] reported PTV $D_{95\%}$ and bladder D_{mean} differences of 0.4% and 1.8%, respectively, utilizing a 3D U-net. Lempart et al.[18] utilized a 2.5D model and achieved differences of 1% and 2.1%. However, these works predicted dose for volumetric modulated arc therapy-based treatments, which differ from the treatment approach in this work. Murakami et al.[23] utilized a GAN-approach resulting in differences of 0.4% and 2% for IMRT-based treatments. Our results of 1.6% and 0.7% compare favorably even with predicted contours as input to the dose prediction network. Some outliers with respect to PTV coverage are however seen with the predicted contour input, where the worst case is an 8 Gy underestimation for $D_{98\%}$. This is due to an underestimation of the CTV volume and

hence the predicted dose does not completely cover the true target volume. This is exemplified in Figure S4.

A few general design choices for the whole pipeline warrant a discussion. First was the choice of utilizing two separate networks. One could have utilized a single network, producing both the segmentation and the dose prediction in a multitask approach. Although this also results in both segmentation and dose prediction as final output, there exists no possibility to intervene in the process, for example, correcting structures before predicting the dose. One also cannot investigate the impact of using predicted instead of true contours as input to the dose prediction. This also does not allow for a potential exchange of a single model, that is, replacing only the segmentation or dose prediction network. One could have utilized the output from the segmentation network as input to the dose prediction during training, possibly gaining a bit of accuracy in the dose prediction where the network could have corrected potential errors from the segmentation. However, this procedure would inescapably connect the two networks, which would disable the possibility of replacing one of them. We also input the PTV instead of the CTV so as not to hard wire the PTV-margins into the dose prediction network. This separation additionally allows for using the dose prediction network for investigating the dosimetric impact of different margin recipes.

One uncertainty in the manual segmentation is both intra- and inter-observer variability. Since the CTV was delineated by different observers (both within and between patients) both variabilities occur for this contour. However, it should be noted that none of the delineations were performed from scratch, but rather from a suggestion following a deformable registration from the reference plan, which may decrease the variability. For the OARs only intra-observer variability is applicable, due to the retrospective correction by a single observer. The remaining contours applied as the input to the dose prediction, that is, the body contour and bone structure, were taken from the treatment data and hence not modified. It should be noted that these contours are generally not modified during treatment but rather just rigidly/deformably propagated onto the daily image and left as is. There are also uncertainties attributed to the ground truth dose matrices. Since the reference dose matrices were obtained from optimization based on varying accuracy on the segmentation of the OARs there is a discrepancy between these, and doses obtained with accurately segmented OARs. Although this was handled by inspection and correction of the worst cases (see Supplementary Material) it is nonetheless present. Additionally, both reference and daily plans are optimized and approved by different people. Although supplied DVCs during the process, slightly different planning strategies (e.g., tweaking of optimization criteria) and approval strategies (e.g., prioritizing between target coverage and OAR sparing) could introduce some noise in the final dose. Also, since the optimization is based on Monte-Carlo dose calculation the noise may induce some variation, that is, two optimizations following each other will produce slightly different plans (see Figure S1).

The mean inference times for the full pipeline of 110 s may not be regarded as purely an idle time in the workflow since, for example, importing of the daily image and initial fusion with the reference image could be performed simultaneously. In this work, we processed the whole acquired patient volume, although a very large part is well beyond the volume of interest. More distinct cropping of the volume prior to all the image processing steps could reduce the inference times significantly. Additionally, inference times are tightly coupled to the hardware, in which higher-performance hardware would decrease the inference times.

An envisioned workflow utilizing such a tool developed here is: Following daily image acquisition, send the MR-images to the online planning system (as current routine) and additionally to this tool. After images have been imported and image fusion performed with the reference scan in the online system, segmentations and a predicted dose for an intended ATS workflow will be ready. Additionally, the reference dose distribution can be rigidly aligned with the daily image to predict an ATP dose distribution (not developed here). Based on these predictions, the ATP or ATS workflow can be chosen. If deciding to go for the ATS workflow, the predicted segmentations may be imported into the online system and the workflow continue. Although an additional step in the pipeline is added with the decision-making process, the time required for this step is likely very short, and could additionally be automated through the comparison of clinical DVCs between the predicted plans.

5 | CONCLUSION

In conclusion, combining a deep learning segmentation and dose prediction network exhibited on average <2% difference between predicted and true dose distributions for all structures and DVCs for daily prostate treatment on the MR-Linac. This could hence be used in the decision-making process in the daily treatment chain where differences are >2%. In addition, overall high DSC values >0.87 indicate the usability of the predicted contours in the clinical workflow.

ACKNOWLEDGMENTS

This work was funded by the project ASSIST (2019-04735) through Vinnova, Sweden's innovation agency.

CONFLICT OF INTEREST STATEMENT

David Tilly was part-time employed by Elekta Instruments AB during the work on this project. David Tilly's involvement in the presented work was solely in the

capacity as an employee at the Uppsala University Hospital. Elekta was not involved in any way other than as a vendor of equipment used in clinical routine at Uppsala University Hospital.

DATA AVAILABILITY STATEMENT

Image data and contours will not be made publicly available.

REFERENCES

1. Winkel D, Bol GH, Kroon PS, et al. Adaptive radiotherapy: the Elekta Unity MR-linac concept. *Clin Transl Radiat Oncol*. 2019;18:54-59. doi:10.1016/j.ctro.2019.04.001
2. Klüter S. Technical design and concept of a 0.35 T MR-Linac. *Clin Transl Radiat Oncol*. 2019;18:98-101. doi:10.1016/j.ctro.2019.04.007
3. Lamb J, Cao M, Kishan A, et al. Online adaptive radiation therapy: implementation of a new process of care. *Cureus*. 2017;9(8):e1618. doi:10.7759/cureus.1618
4. de Muinck Keizer DM, Kerkmeijer LGW, Willigenburg T, et al. Prostate intrafraction motion during the preparation and delivery of MR-guided radiotherapy sessions on a 1.5T MR-Linac. *Radiother Oncol*. 2020;151:88-94. doi:10.1016/j.radonc.2020.06.044
5. Schaule J, Chamberlain M, Wilke L, et al. Intrafractional stability of MR-guided online adaptive SBRT for prostate cancer. *Radiat Oncol*. 2021;16(1):189. doi:10.1186/s13014-021-01916-0
6. Lim SY, Tran A, Tran ANK, et al. Dose accumulation of daily adaptive plans to decide optimal plan adaptation strategy for head-and-neck patients treated with MR-Linac. *Med Dosim*. 2022;47(1):103-109. doi:10.1016/j.meddos.2021.08.005
7. Xia WL, Liang B, Men K, et al. Prediction of adaptive strategies based on deformation vector field features for MR-guided adaptive radiotherapy of prostate cancer. *Med Phys*. 2023;50(6):3573-3583. doi:10.1002/mp.16192
8. Bohoudi O, Bruynzeel AME, Meijerink MR, et al. Identification of patients with locally advanced pancreatic cancer benefitting from plan adaptation in MR-guided radiation therapy. *Radiother Oncol*. 2019;132:16-22. doi:10.1016/j.radonc.2018.11.019
9. Parchur AK, Lim S, Nasief HG, et al. Auto-detection of necessity for MRI-guided online adaptive replanning using a machine learning classifier. *Med Phys*. 2023;50(1):440-448. doi:10.1002/mp.16047
10. La Macchia M, Fellin F, Amichetti M, et al. Systematic evaluation of three different commercial software solutions for automatic segmentation for adaptive therapy in head-and-neck, prostate and pleural cancer. *Radiat Oncol*. 2012;7(1):1-16. doi:10.1186/1748-717X-7-160
11. Doolan PJ, Charalambous S, Roussakis Y, et al. A clinical evaluation of the performance of five commercial artificial intelligence contouring systems for radiotherapy. *Front Oncol*. 2023;13(August):1-13. doi:10.3389/fonc.2023.1213068
12. Lustberg T, van Soest J, Gooding M, et al. Clinical evaluation of atlas and deep learning based automatic contouring for lung cancer. *Radiother Oncol*. 2018;126(2):312-317. doi:10.1016/j.radonc.2017.11.012
13. Boulanger M, Nunes JC, Chourak H, et al. Deep learning methods to generate synthetic CT from MRI in radiotherapy: a literature review. *Physica Med*. 2021;89(July):265-281. doi:10.1016/j.ejmp.2021.07.027
14. Poel R, Kamath AJ, Willmann J, et al. Deep-learning-based dose predictor for glioblastoma–assessing the sensitivity and robustness for dose awareness in contouring. *Cancers (Basel)*. 2023;15(17):4226. doi:10.3390/cancers15174226
15. Kamath A, Poel R, Willmann J, Andratschke N, Reyes M. How sensitive are deep learning based radiotherapy dose prediction models to variability in organs at risk segmentation? *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, Cartagena, Colombia, 2023:1-4. doi:10.1109/isbi53787.2023.10230559
16. Nguyen D, Long T, Jia X, et al. A feasibility study for predicting optimal radiation therapy dose distributions of prostate cancer patients from patient anatomy using deep learning. *Sci Rep*. 2019;9(1):1-10. doi:10.1038/s41598-018-37741-x
17. Ronneberger O, Fischer P, Brox T. U-net: convolutional networks for biomedical image segmentation. *Med Image Comput Comput Assist Interv*. 2015;9351:234-241. doi:arXiv:1505.04597.
18. Lempart M, Benedek H, Nilsson M, et al. Volumetric modulated arc therapy dose prediction and deliverable treatment plan generation for prostate cancer patients using a densely connected deep learning model. *Phys Imaging Radiat Oncol*. 2021;19(May):112-119. doi:10.1016/j.phro.2021.07.008
19. Ma M, Kovalchuk N, Buyyounouski MK, Xing L, Yang Y. Incorporating dosimetric features into the prediction of 3D VMAT dose distributions using deep convolutional neural network. *Phys Med Biol*. 2019;64(12):125017. doi:10.1088/1361-6560/ab2146
20. Kandalan RN, Nguyen D, Rezaeian NH, et al. Dose prediction with deep learning for prostate cancer radiation therapy: model adaptation to different treatment planning practices. *Radiother Oncol*. 2020;153:228-235. doi:10.1016/j.radonc.2020.10.027
21. Mashayekhi M, Tapia IR, Balagopal A, et al. Site-agnostic 3D dose distribution prediction with deep learning neural networks. *Med Phys*. 2022;49(3):1391-1406. doi:10.1002/mp.15461
22. Osman AFI, Tamam NM. Attention-aware 3D U-Net convolutional neural network for knowledge-based planning 3D dose distribution prediction of head-and-neck cancer. *J Appl Clin Med Phys*. 2022;23(7):e13630. doi:10.1002/acm2.13630
23. Murakami Y, Magome T, Matsumoto K, Sato T, Yoshioka Y, Oguchi M. Fully automated dose prediction using generative adversarial networks in prostate cancer patients. *PLoS One*. 2020;15(5):1-16. doi:10.1371/journal.pone.0232697
24. Li H, Peng X, Zeng J, et al. Explainable attention guided adversarial deep network for 3D radiotherapy dose distribution prediction. *Knowl Based Syst*. 2022;241:108324. doi:10.1016/j.knosys.2022.108324
25. Thomas MA, Fu Y, Yang D. Development and evaluation of machine learning models for voxel dose predictions in online adaptive magnetic resonance guided radiation therapy. *J Appl Clin Med Phys*. 2020;21(7):60-69. doi:10.1002/acm2.12884
26. Dice LR. Measures of the amount of ecologic association between species. *Ecology*. 1945;26:297-302. doi:10.2307/1932409
27. Phil T, Albrecht T, Gay S, Rasmusse ME. Sikerdebaard/dcmrtstruct2nii: dcmrtstruct2nii v5 (Version v5). doi:10.5281/zenodo.4037864
28. Yaniv Z, Lowekamp BC, Johnson HJ, Beare R. SimpleITK image-analysis notebooks: a collaborative environment for education and reproducible research. *J Digit Imaging*. 2018;31(3):290-303. doi:10.1007/s10278-017-0037-8
29. Beare R, Lowekamp B, Yaniv Z. Image segmentation, registration and characterization in R with simpleITK. *J Stat Softw*. 2018;86(8):1-35. doi:10.18637/jss.v086.i08
30. Lowekamp BC, Chen DT, Ibáñez L, Blezek D. The design of simpleITK. *Front Neuroinform*. 2013;7(DEC):1-14. doi:10.3389/fninf.2013.00045
31. Oktay O, Schlemper J, Folgoc LL, et al. Attention U-Net: learning where to look for the pancreas. arXiv. 2018. arXiv:1804.03999.
32. Isensee F, Jaeger PF, Kohl SAA, Petersen J, Maier-Hein KH. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nat Methods*. 2021;18(2):203-211. doi:10.1038/s41592-020-01008-z

33. Ulyanov D, Vedaldi A, Lempitsky V. Instance normalization: the missing ingredient for fast stylization. *arXiv*. 2017. arXiv:1607.08022.
34. Kingma DP, Ba JL. Adam: a method for stochastic optimization. *Proceedings of the 3rd International Conference on Learning Representations*, San Diego, CA, USA, 7–9 May 2015:1-15.
35. Pielawski N, Wählby C. Introducing Hann windows for reducing edge-effects in patch-based image segmentation. *PLoS One*. 2020;15(3):e0229839. doi:10.1371/journal.pone.0229839
36. Jhanwar G, Dahiya N, Ghahremani P, Zarepisheh M, Nadeem S. Domain knowledge driven 3D dose prediction using moment-based loss function. *Phys Med Biol*. 2022;67(18):185017. doi:10.1088/1361-6560/ac8d45

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Fransson S, Strand R, Tilly D. Deep learning-based dose prediction for magnetic resonance-guided prostate radiotherapy. *Med Phys*. 2024;51:8087–8095. https://doi.org/10.1002/mp.17312